
This is the **published version** of the bachelor thesis:

Lendínez Gutiérrez, Alberto; Llisteri Boix, Joaquim, dir. Assignació automàtica de signes de puntuació a partir de les pauses en reconeixement de la parla. 2016. 39 pag. (835 Grau en Estudis d'Anglès i Català)

This version is available at <https://ddd.uab.cat/record/170488>

under the terms of the  license

Assignació automàtica de signes de puntuació a partir de les pauses en reconeixement de la parla



**Universitat Autònoma
de Barcelona**

A handwritten signature in black ink, which appears to read 'Alberto Lendínez Gutiérrez', is written over a horizontal line.

Autor: Alberto Lendínez Gutiérrez
Grau en estudis d'Anglès i Català
Tutor: Joaquim Llisterri Boix
Curs: 2016 – 2017

AGRAÏMENTS

Primerament, voldria agrair al meu tutor Joaquim Llisterri la seva disponibilitat horària i paciència a l'hora de guiar-me durant la redacció d'aquest treball. També voldria agrair a les coordinadores de l'assignatura (Montserrat Amores, María Jesús Machuca i Marta Oller) que s'hagi pogut realitzar la presentació i qualificació del present treball al primer semestre per tal d'agilitzar la meva graduació. Agraïments especials (com no podria ser d'altra manera) a l'Alberto Acero per l'enregistrament del material acústic que s'ha fet servir com a objecte d'estudi a la part pràctica, així com també als voluntaris que van fer possible la gravació de les frases. Moltes gràcies a la Meritxell Andreu, per ajudar-me en la revisió dels models matemàtics emprats per arribar a l'obtenció dels resultats pertinents.

Finalment, desitjaria esmentar la contínua força amb la que els meus pares i la meva àvia han lluitat davant les adversitats de la vida. Aquesta fortalesa m'ha permès inspirar-me per tal d'aconseguir els meus propòsits al llarg de tota la meva vida estudiantil, tant a l'enginyeria com a la filologia. Mai no els podré agrair amb paraules ni fets tot el que han fet per mi.

ÍNDIX

1. INTRODUCCIÓ	4
2. EL RECONeixEMENT AUTOMÀTIC DE LA PARLA	5
2.1 Estratègies per al reconeixement automàtic de la parla	6
2.1.1 Fase d'entrenament del sistema.....	6
2.1.2 Reconeixement de paraules aïllades	7
2.1.3 Descodificació acústico-fonètica.....	8
2.1.4 Reconeixement de la parla contínua	8
2.2 Aplicacions del reconeixement automàtic de la parla	10
2.2.1 Aplicacions relacionades amb el control de sistemes.....	10
2.2.2 Aplicacions relacionades amb la creació i recreació de text.....	11
3. LES PAUSES I LA PUNTUACIÓ	13
3.1 Tipus i durada de les pauses	13
3.2 Els signes de puntuació	14
3.3 Relació entre pausa i puntuació	15
4. ESTUDI DE LES PAUSES SILENCIOSES.....	16
4.1 Objectius	16
4.2 Metodologia	17
4.2.1 Corpus.....	17
4.2.2 Locutors	17
4.2.3 Gravació	18
4.2.4 Procediment d'anàlisi	18
4.3 Resultats.....	18
5. CONCLUSIONS.....	22
REFERÈNCIES BIBLIOGRÀFIQUES.....	24
ANNEX 2: Resultats per locutor	26
ANNEX 2: Corpus i contexts d'aparició de signes de puntuació.....	32
ANNEX 3: Codis de <i>MatLab</i> per a la realització de les figures	37

1. INTRODUCCIÓ

Les tecnologies del llenguatge (TL o LT, de l'anglès *Language Technologies*) són una disciplina que sorgeix com a resposta a les necessitats d'interacció oral o escrita entre l'ésser humà i els sistemes informàtics per tal d'establir una comunicació bidireccional el més senzilla i satisfactòria possible. Entre les aplicacions de les TL es troben el tractament de textos escrits (correcció de l'ortografia o traducció automàtica, per exemple), així com el reconeixement i la generació de la parla (dictat automàtic o conversió de text en parla). La seva primerenca aparició –les TL van sorgir a finals de la dècada dels 60 (Miñana, 2014: p. 86)–, les anomalies funcionals i la falta de robustesa de segons quines aplicacions fan que les tecnologies del llenguatge siguin una disciplina en la qual encara queda molt per experimentar i millorar. En el marc de les TL es troba una branca coneguda amb el nom de tecnologies de la parla (TP). Segons la definició de Rodríguez Fonollosa (2010),

La tecnologia de la parla estudia com utilitzar la veu, la forma més habitual de comunicar-nos entre persones, per facilitar l'ús i la interacció amb les màquines, ja sigui per si sola o en combinació amb altres dispositius com el teclat o les pantalles tàctils (p. 1).

Tot i així, es podria dir que les TP encara no han arribat a resultats satisfactoris en el tractament dels elements suprasegmentals, com, per exemple, en l'assignació de l'entonació adequada en la conversió de text en parla, en la identificació de l'accent característic d'un parlant o en el mètodes per a assignar de manera automàtica els signes de puntuació a partir de les pauses en el reconeixement.

Sense cap tipus de pretensió, aquest treball provarà d'aportar solucions al problema de l'assignació de signes de puntuació en el reconeixement automàtic de la parla, ja que es considera pertinent i de relativa importància assolir un mètode que permeti introduir els signes de puntuació en el text de sortida d'un reconeixedor sense que l'usuari els hagi de dictar. Aquesta possibilitat, per exemple, marcaria un abans i un després a l'àmbit del subtitulat automàtic (vegeu l'apartat 2.2.2), ja que avui dia encara no s'ha pogut arribar a una solució (ni tan sols, pel que sabem, a desenvolupar una aplicació) per aquest problema relacionat amb la generació d'un text a partir d'un senyal sonor.

2. EL RECONeixEMENT AUTOMÀTIC DE LA PARLA

El reconeixement automàtic de la parla (RAP, o ASR, de l'anglès *Automatic Speech Recognition*) tracta de permetre la interacció oral entre éssers humans i sistemes computacionals. El seu funcionament es fomenta principalment en el processament d'un senyal acústic que és produït per l'usuari. L'aplicació haurà d'extreure'n les característiques i descodificar la informació que conté el missatge, fet que es realitzarà mitjançant unes tècniques basades en procediments d'aprenentatge automàtic. A la figura 1 es presenten, de manera esquematitzada, els principal mòduls d'un sistema de RAP.

En el mòdul de processament de senyal es tracta el senyal sonor mitjançant un equalitzador¹ i els segments d'àudio entrants es separen en funció de si es consideren informació sonora o soroll. Un cop realitzat el tractament, els segments d'àudio considerats com a “bons”, és a dir, com a part d'un senyal de parla, s'envien al mòdul d'extracció de característiques, en el qual es realitza un procés de parametrització (Méndez, Nadeu, Hernando i Vallverdú, 1995) que permet extreuen tot aquells valors i paràmetres distintius (coeficients LPC, característiques temporals, etc.) que conformen tot un conjunt de trets significatius anomenats “vectors de característiques” (Duxans i Ruiz, 2012: p. 10). Aquests vectors seran descodificats al darrer mòdul que es mostra a la figura 1 i comparats de manera estadística amb els models acústics i els models de llenguatge que s'han creat durant la fase d'entrenament del sistema (vegeu l'apartat 2.1.1) per tal d'arribar a una transcripció automàtica satisfactòria.

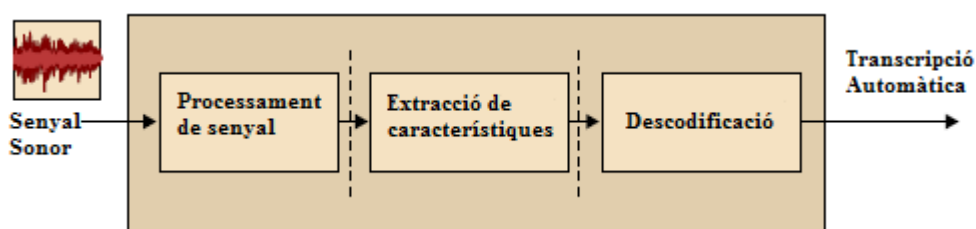


Figura 1. Diagrama de blocs del funcionament d'un sistema de RAP (Duxans i Ruiz, 2012, p. 10).

¹ Un equalitzador és un instrument que ajusta la intensitat de les freqüències de l'espectre sonor per a compensar deficiències acústiques en segons quina freqüència o per emfatitzar-ne d'altres.

2.1 Estratègies per al reconeixement automàtic de la parla

Quan les tecnologies de la parla van sorgir es feia servir una tècnica per al reconeixement automàtic basada, únicament, en mòduls de comparació entre unitats emmagatzemades (Martí *et al.*, capítol 2, p. 46-47) però, degut a la seva baixa funcionalitat, aquesta tècnica va quedar obsoleta i es va substituir per les tres tècniques o estratègies que actualment conformen els sistemes de reconeixement automàtic de la parla: el reconeixement de paraules aïllades, la descodificació acústico-fonètica i el reconeixement de la parla contínua. Abans de parlar de les tècniques, però, s'introduirà breument el mòdul d'entrenament del sistema per tal de facilitar la comprensió dels conceptes que seran tractats amb posterioritat.

2.1.1 Fase d'entrenament del sistema

La fase d'entrenament del sistema és totalment necessària a qualsevol aplicació de reconeixement de la parla ja que, sense aquesta etapa, cap aplicació seria capaç de funcionar correctament. El principal element d'aquesta fase és el corpus dissenyat per a l'entrenament del sistema: un conjunt ampli i estructurat d'exemples reals d'ús de la llengua, ja siguin texts o mostres orals, que són emmagatzemats a una base de dades electrònica (Kennedy, 1998) i que es faran servir posteriorment per comparar el senyal d'entrada amb els models emmagatzemats (Moure i Llisterri, 1996). Amb aquests elements, es procedeix a la creació d'unes "plantilles" de referència que permetin la comparació amb els enuncisats emesos per l'usuari. Abans, però, s'haurà de delimitar quines seran les unitats amb les que treballarà el sistema i que, en el nivell acústic, consisteixen en segments sonors que inclouen la transició entre dos o més fons consecutius, anomenats difonemes si inclouen dos elements, trifonemes si n'incorporen tres, etc. Aquestes plantilles, a més, hauran d'estar parametritzades (vegeu l'apartat 2) segons la seva estructura acústica (Moure i Llisterri, 1996).

Un cop finalitzada la creació de les plantilles de referència, el sistema estarà preparat per al reconeixement de nous enuncisats.

2.1.2 Reconeixement de paraules aïllades

El reconeixement de paraules aïllades es basa principalment en la comparació del senyal sonor emès per l'usuari amb uns models de referència. Mitjançant una digitalització del senyal acústic i el tractament explicat a l'apartat 2, el sistema serà capaç de processar el senyal refinat² i distingir els principis i finals de cada paraula. Després d'aquest procés, l'aplicació extreu els vectors de característiques pertinents i els compara amb els models de referència creats durant l'etapa d'entrenament mitjançant uns models estadístics coneguts com a “models ocults de Markov” (HMM, de l'anglès *Hidden Markov Models*)³. Els HMM engloben les propietats, les cadenes i els models de Markov, models que encara avui dia continuen vigents en el camp de l'aprenentatge automàtic (Juang i Rabiner, 1993), ja que els paradigmes d'aprenentatge automàtic que es fan servir en la creació de xarxes d'informació estan basats en xarxes neuronals, però matemàticament segueixen els principis establerts pels models de Markov. Aquests models es basen en el tractament de magnituds aleatòries que varien en funció del temps i d'una altra variable correlativa, la qual pot ser directament observable però no predictable (Rabiner, 1989). A la figura 2 es pot veure exemplificat de manera gràfica un model de Markov senzill.

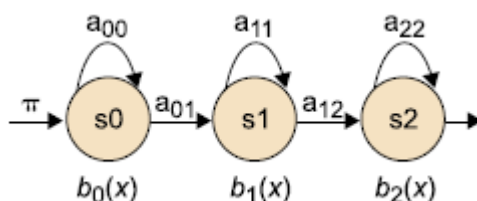


Figura 2. Exemple d'un model ocult de Markov de tres estats (Duxans i Ruiz, 2012: p. 19).

A més, els models estadístics necessaris per establir les comparacions entre els vectors de característiques i els diccionaris (vegeu l'apartat 2.1.1) o models de referència hauran de complir la propietat de Markov⁴. En l'última etapa d'un sistema de reconeixement de paraules aïllades, es compararà la distància entre el senyal d'entrada emès per l'usuari i

² S'entén com “refinat” en aquest context el resultat d'un procés que ha canviat el senyal inicial. En aquest cas, el senyal refinat és l'equivalent al senyal inicial digitalitzat i parametritzat.

³ Andrei Andreièvitx Màrkov (1856-1922) fou matemàtic creador de quantioses teories sobre els nombres i la probabilitat, entre les quals es troben les de variables aleatòries i independents que conformen el que es coneix en RAP com a “models ocults de Markov”.

⁴ La propietat de Markov il·lustra que el valor d'una variable aleatòria depèn únicament del seu valor present, i no del procediment que s'ha fet servir per arribar a tal valor.

el model de referència creat durant l'entrenament del sistema per tal d'arribar a la transcripció de les paraules corresponents.

2.1.3 Descodificació acústico-fonètica

La descodificació acústico-fonètica (DAF) és una tècnica imprescindible en aplicacions orientades al reconeixement de vocabulari molt ampli. El senyal acústic emès per l'usuari serà analitzat per tal d'extreure els paràmetres més importants de cada element (Lea, 1986). Després, es procedeix a la segmentació del senyal d'entrada en elements fonètics més petits que la paraula coneguts amb el nom d'unitats sublèxiques (Moure i Llisterra, 1996). Aquests segments seran classificats fonèticament i es compararan amb els diccionaris de pronúncia (*pronunciation lexicon*) del sistema per tal de generar una primera transcripció aproximada de l'enunciat a partir de les coincidències establertes. Per arribar a una resolució satisfactòria de la transcripció, el sistema haurà d'ajudar-se d'un model de llenguatge (Oostdijk i de Haan, 1994), el qual consisteix en un “conjunto de probabilidades de transición entre palabras” (Moure i Llisterra, 1996: p. 84). Aquest model de llenguatge s'obté a partir del corpus textual emprat per a l'entrenament (vegeu l'apartat 2.1.1) i és capaç de millorar la transcripció final de l'enunciat, ja que es descarten hipòtesis generades per la comparació amb els models acústics gràcies a la informació sobre la probabilitat que una paraula aparegui en un context determinat.

El reconeixement de grans vocabularis no només requereix l'anàlisi acústica i fonètica: els analitzadors sintàctics (en anglès, *parsers*) també són necessaris per a completar la descodificació acústico-fonètica, així com per a incorporar al reconeixement informació prosòdica com la relativa als moviments i les interrupcions de la corba melòdica (Waibel, 1986).

2.1.4 Reconeixement de la parla contínua

Les estratègies per al reconeixement de la parla contínua són les més complexes i requereixen la utilització de tècniques com les que s'emprenen en la DAF. Per a poder dur a terme el reconeixement de la parla contínua, el sistema ha de contenir un model de llenguatge (vegeu l'apartat 2.1.3) amb seqüències de n elements (n -grames) extretes d'un corpus textual (Baker, Hardie, i McEnery, 2006). A més a més, els elements han de ser consecutius i han d'estar definit pels HMM. Els models de referència que farà

servir el sistema hauran d'incloure també una representació fonètica del lèxic, en forma del que es sol anomenar un diccionari de pronunciació (*pronunciation lexicon*). El sistema segmentarà en difonemes (vegeu l'apartat 2.1.1) el senyal refinat emès per l'usuari i el compararà amb els models de referència esmentats amb anterioritat. Després, mitjançant algorismes com el de Baum-Welch⁵, es trobarà quin és el vector de característiques que té una probabilitat més alta de fer que el resultat sigui el més semblant possible al senyal d'entrada. L'algorisme de Baum-Welch es fa servir a reconeixement de la parla contínua perquè les dades del senyal emès per l'usuari contenen molta més informació que en el cas del reconeixement de paraules aïllades, i tant els HMM com els models de referència han de ser comparats contínuament entre si. Com a darrer pas, el sistema cercarà el resultat que més s'apropi estadísticament a la solució que consideri convenient i establirà la transcripció més adient per a l'enunciat que estigui tractant. A la figura 3 es pot observar un exemple visual d'un model de RAP.

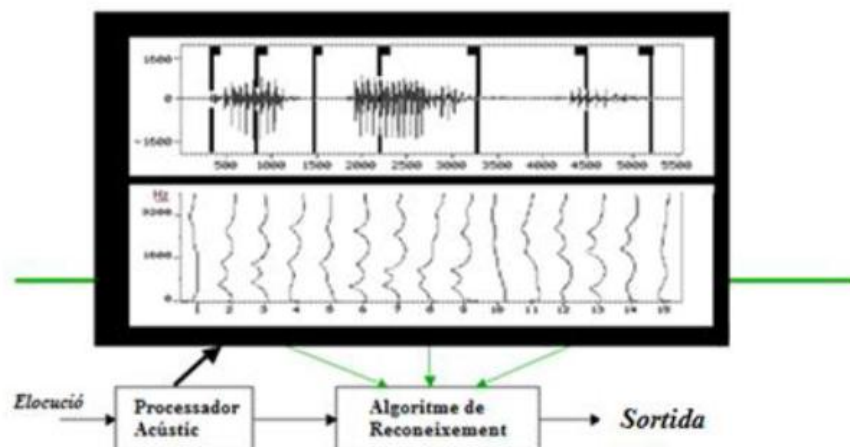


Figura 3. Exemple d'un model de reconeixement de parla contínua (Adaptada de Mariño, s.d.).

⁵ L'algorisme de Baum-Welch facilita trobar els paràmetres X que maximitzen la probabilitat d'un HMM mitjançant un procés iteratiu. L'algorisme calcula les probabilitats i crea un nou model amb una probabilitat d'èxit més gran que l'anterior.

2.2 Aplicacions del reconeixement automàtic de la parla

Existeixen diverses aplicacions per als sistemes de RAP i, si bé és cert que abasten un ampli espectre d'usos, generalment tots els dissenys es basen en les estratègies descrites a l'apartat 2.1. Donada la gran quantitat d'aplicacions possible, es distingirà entre els dos àmbits en els que les aplicacions són més comunes, i es realitzarà una breu explicació d'aquestes aplicacions.

2.2.1 Aplicacions relacionades amb el control de sistemes

Es troben, en primer lloc, aplicacions del RAP orientades a controlar sistemes electrònics i industrials mitjançant ordres verbals. Aquests sistemes treballen, bàsicament, amb paraules aïllades o bé amb enunciats curts i previsibles i, tot i que tenen un gran nombre d'aplicacions potencials de cara a un futur pròxim, se'n farà només una breu descripció, ja que no constitueixen l'objectiu del treball. Entre aquestes aplicacions es poden assenyalar les següents:

- Serveis interactius telefònics. Les empreses que ofereixen un servei determinat o bé les que han de disposar d'un servei tècnic per cobrir les necessitats de l'usuari són les més beneficiades amb aquesta aplicació, si bé és cert que, actualment, en aquest darrer cas la part automàtica del servei telefònic és limitada i serveix d'enllaç entre l'usuari i el tècnic de manteniment.
- Sistemes mòbils amb interfície interactiva i reconeixement de parla. Es tracta, en aquest cas, de tauletes tàctils (*tablets*) o telèfons mòbils que obren una determinada aplicació després que l'usuari n'hagi pronunciat el nom en veu alta. En aquesta categoria també s'inclouen els GPS i els navegadors d'Internet que poden realitzar la cerca de direccions o de dades que l'usuari hagi demanat.
- Robòtica i electrònica. El reconeixement de paraules aïllades es pot fer servir amb robots i màquines que estiguin preparats per rebre ordres de manera oral. Aquest tipus de reconeixement és també molt útil en el control i l'automatització de processos industrials.
- Sistemes d'ajuda a discapacitats com, per exemple, aplicacions de domòtica.

Encara queden aplicacions possibles per esmentar, i de ben segur que es descobriran noves possibilitats a mesura que s'avanci la recerca i es millorin els sistemes de reconeixement automàtic de la parla en relació amb aspectes de l'enginyeria industrial.

2.2.2 Aplicacions relacionades amb la creació i recreació de text

Els sistemes relacionats amb la creació de textos estan més centrats en el tractament de la parla contínua i de la parla espontània, i s'empren per al dictat de textos o per a la transcripció del discurs oral. Aquestes aplicacions són les que presenten més interès per al present treball. Entre les més conegudes es poden destacar:

- Sistemes de dictat automàtic. El dictat automàtic és una de les aplicacions més funcionals i útils per facilitar la feina a professions per als quals el redactat de documents és una necessitat diària (per exemple, en els àmbits mèdic o jurídic). També tenen especial interès les aplicacions de missatgeria en sistemes mòbils (WhatsApp seria un exemple conegut per tothom), així com el dictat de correus electrònics en aparells amb teclats petits.
- Creació automàtica de subtítols. Els portals d'Internet amb elements audiovisuals (YouTube, per exemple) són els més afavorits per la possibilitat de crear automàticament subtítols. A partir d'un discurs oral, un sistema de reconeixement automàtic tractarà de processar l'enregistrament que es consulta -mitjançant les tècniques descrites a l'apartat 2.1.1- per tal de presentar el missatge corresponent per escrit. El principal problema es troba en el tractament dels elements segmentals i suprasegmentals propis del parlant, ja que en gran mesura reflecteixen característiques individuals. També es pot afegir una altra possibilitat emprant eines pròpies del processament del llenguatge natural: la traducció de subtítols amb tècniques de traducció automàtica.

Cal destacar que només el dictat automàtic necessita del dictat dels signes de puntuació per part de l'usuari, que ha d'indicar quin tipus de signe de puntuació vol afegir en cada moment. Això pot resultar problemàtic a l'hora de realitzar una conversió de parla en text de manera satisfactòria, ja que el fet que l'usuari pronunciï, per exemple, la paraula “coma” pot donar lloc a dues possibilitats: o bé que el sistema inclogui el signe de puntuació corresponent a una coma o, pel contrari, que escrigui la pròpia paraula “coma” al text. Per tant, els sistemes descrits en aquest apartat comparteixen un mateix

problema potencial: la generació automàtica de signes de puntuació. A la figura 4 es troba un exemple de subtitulat automàtic d'un vídeo de YouTube que il·lustra a la perfecció les mancances en aquest tipus d'aplicacions. Tot i que el subtitulat reproduceix perfectament el discurs oral produït pel presentador de l'informatiu, transcrivint gairebé amb una precisió total totes les paraules pronunciades, l'absència de signes de puntuació fa que la separació entre frases sigui més difícil d'establir pel receptor del missatge tot i que, en aquest exemple en concret, els punts i apart han estat substituïts per canvis de paràgraf, cosa que facilita la comprensió. En canvi, es pot observar que, en comptes de les cometes que haurien d'obrir una citació, apareix la paraula "quote" pronunciada pel presentador de l'informatiu.

La generació automàtica de signes de puntuació, poc valorada a sectors més orientats cap a l'electrònica i la indústria, es converteix en una condició de primera necessitat si es vol disposar d'un text correctament puntuat partint d'un discurs oral emès de manera natural per l'usuari. Les principals responsables de l'addició o no de signes de puntuació són elements conformatos per silencis o allargaments vocàlics de duració determinada i que delimiten un grup fònic⁶ o una oració (Gil, 2007: p. 56). S'està fent referència, com és evident, a les pauses.



Figura 4. Exemple de subtitulat automàtic en anglès. (Arirang News, 26 d'Octubre de 2016).

⁶ Segons Gil (2007: p. 56): "Fragmento de discurso comprendido entre dos pausas sucesivas. Algunos autores lo denominan *grupo de espiración o espiratorio*".

3. LES PAUSES I LA PUNTUACIÓ

En un discurs oral, un ús adequat de les pauses i dels moviments de la freqüència fonamental que precedeixen i segueixen la seva aparició determinen una divisió en constituents que permet transmetre aquells conceptes considerats com a importants, facilitant així la correcta comprensió del discurs. En un text escrit, els signes de puntuació tenen també com a funció segmentar i organitzar la informació. En aquest apartat es definiran els conceptes de pausa i de signe de puntuació, així com també les seves característiques bàsiques i la seva utilització normativa. Finalment, s'explicarà la relació entre les pauses i els signes de puntuació en el discurs oral i en l'escrit.

Començant amb les explicacions conceptuals, les pauses es podrien caracteritzar com “elementos constitutivos del ritmo en el lenguaje oral” (Rebollo, 1997: p. 2). Una altra definició possible seria la proporcionada per Gil (2007), la qual descriu la pausa com “un silencio o vocalización intercalados en el discurso” (p.544). Les pauses tenen un paper molt important al discurs oral i seran un element decisiu a l'hora d'introduir els signes de puntuació en les aplicacions de les tecnologies de la parla relacionades amb la creació i recreació de text (vegeu l'apartat 3.3).

3.1 Tipus i durada de les pauses

En el discurs oral existeixen dos tipus de pauses que es distingeixen segons si es tracta de pauses sonores o “plenes” o bé de pauses silencioses o “buides”; aquest treball es centrarà principalment en les pauses silencioses, és a dir, aquelles en les quals es produeix una interrupció de l'energia sonora. Les característiques acústiques com la durada són diferents segons si es tracta d'un o altre tipus de pausa. Per tant, s'hauran de tenir en compte aquests fets a l'hora del seu estudi en el discurs oral.

Les pauses sonores es defineixen com a “sonidos o ruidos que salen por la boca, que no son palabras pero que desempeñan funciones comunicativas importantes” (Calsamiglia i Tusón, 1999: p. 7). En el seu estudi, Rebollo (1997) explica que la durada de les pauses sonores en espanyol d'Amèrica varia entre els 53 i els 852 ms. D'altra banda, Brennan i Williams (1995) assenyalen que la durada de les pauses plenes en anglès oscil·la entre els 344 i els 527 ms. Cal esmentar, però, que aquests autors només van estudiar un tipus

de recurs consonàntic i un de vocàlic: l'allargament de la consonant “m” i l'ús de l'interjecció “uh” com a pausa sonora.

Les pauses silencioses, d'altra banda, es defineixen com a intervals silenciosos que interrompen el discurs (Stejskal, 2010). Una definició més centrada en les propietats acústiques és la de Duez (1982), que descriu les pauses silencioses com “qualsevol interval a l'oscil·lograma on l'amplitud del senyal no es pot diferenciar del so ambient” (p. 11. *Traducció: Alberto Lendínez*). Butterworth (1980) considera una pausa silenciosa qualsevol interval de silenci que sigui igual o superior als 200 ms. Campione i Véronis (2002) van estudiar prop de 6000 pauses en cinc llengües diferents i van suggerir que les pauses silencioses es podien diferenciar en funció de la seva durada: pauses breus (inferiors als 200 ms), pauses de durada mitjana (entre els 200 i els 2000 ms) i pauses de durada llarga (més de 2000 ms). Duez també va establir un rang temporal per a la durada de les pauses silencioses en francès comprès entre els 180 i els 250 ms (p. 13), valors que disten força de l'interval comprès entre 42 i 1134 ms. proposat per Rebollo (1997).

En el cas de la llengua catalana no s'han trobat estudis que presentin valors de duració de pauses, però es consideraran com a pauses silencioses els intervals sense energia sonora amb valors superiors als 200 ms, ja que superen els valors propis d'una consonant oclusiva i s'ajusten als rangs de durada de pauses breus i mitjanes proposats per Campione i Véronis (2002). A més, els valors de Duez per a la llengua francesa també s'aproximen força a aquest llinar.

Tant les pauses sonores com silencioses es troben presents en tot tipus de discurs oral, però la seva duració difereix força (com s'ha pogut establir anteriorment) en funció de l'estil de parla. En l'estudi realitzat per Albalá *et al.* (2008), per exemple, es mostra que les pauses sonores i silencioses es comporten de manera diferent segons si apareixen en un discurs oral produït de manera espontània o, pel contrari, si es tracta de la lectura d'un text.

3.2 Els signes de puntuació

Els signes de puntuació són una eina que es fa servir en l'escriptura amb una doble funció: primerament, fer que el propi text sigui comprensible i correctament interpretable pel lector i, d'altra banda, transmetre al lector els matisos propis que el

mateix text posseiria en la seva realització oral. Tot i que hom disposa d'una àmplia gamma de signes de puntuació, el present apartat es centrarà en aquells que, a part de ser el més comuns i utilitzats en la redacció de texts, seran objecte d'estudi a la part experimental: el punt i la coma.

Segons el manual de la Universitat Autònoma de Barcelona *Els signes de puntuació* (1995), la coma (,) “representa una pausa breu que sovint es reflecteix oralment en entonació” (p. 25). Entre els seus usos en la redacció de texts, els considerats més significatius per al present estudi són els següents:

- Separar diferents elements d'una enumeració (excepte quan apareixen les conjuncions *o*, *ni* o *i*, que serveixen d'enllaç entre el penúltim i el darrer element): *No tenim pomes, llimones ni pinya.*
- Delimitar els incisos en el cas de les oracions relatives explicatives: *La veïna, que té una casa a Madrid, vivia al barri de Villaverde.*
- Establir una separació entre elements amb contingut conjuntiu i adverbial (com poden ser *però*, *no obstant això*...) i la resta del text.

El punt (.) correspon a una “pausa forta i llarga que indica que la frase té un significat complet i gramaticalment independent” (UAB, 1995: p. 42). Conseqüentment amb aquesta definició, la seva funció general serà marcar la separació entre dues frases segons si mantenen una relació entre elles (punt i seguit) o si tracten aspectes diferents (punt i apart). Finalment, es posa un punt i final per assenyalar l'acabament del text.

3.3 Relació entre pausa i puntuació

Les definicions proporcionades a l'apartat 3.2, permeten establir que existeix una relació molt estreta entre les pauses i els signes de puntuació (en aquest cas, el punt i la coma). Tant el punt com la coma representen de manera gràfica una pausa de més o menys durada al discurs oral. Així, és lícit afirmar que els signes de puntuació delimiten canvis en la freqüència fonamental que precedirà a una pausa. Aquest fet caldrà considerar-lo en futures recerques realitzades a partir de l'experiment que es presentarà a l'apartat 4.

4. ESTUDI DE LES PAUSES SILENCIOSES

Donat que no s'han pogut localitzar estudis que tractin la durada de les pauses en llengua catalana en el context de les tecnologies de la parla, s'ha considerat adient realitzar una anàlisi de les pauses silencioses en l'àmbit de la lectura de textos. Són molts els treballs sobre les pauses silencioses que s'han dut a terme en diferents llengües, però sempre s'ha tractat d'arribar a un patró general dels valors de les pauses buides. Encara que aquests estudis han proporcionat nombroses dades que permeten avançar en el coneixement del fenomen, es considera oportú canviar l'enfocament dels experiments per tal d'arribar a nous resultats que puguin aportar solucions als problemes relacionats amb les pauses silencioses en les aplicacions de creació i recreació de text.

Una de les principals limitacions de tipus lingüístic és que la durada de cada signe de puntuació no està normalitzada i, per tant, cada parlant realitzarà pauses més o menys llargues segons el seu criteri. Així doncs, els valors que s'obtinguin de l'anàlisi d'un corpus estaran condicionats pels locutors. Per aquesta raó, l'estudi que es realitzarà a continuació seguirà una metodologia de caire matemàtic funcional. Mitjançant aquest enfocament es pretén trobar (si existeix) una correlació matemàtica entre el corpus que es tractarà a l'experiment i els propis parlants.

4.1 Objectius

L'objectiu principal de l'experiment és demostrar l'existència d'una correlació matemàtica entre la durada dels signes de puntuació, el seu context (l'aparició del signe de puntuació abans i després de vocal o abans de vocal i després de consonant) i els diferents parlants. Aquesta correlació facilitaria definir una equació característica de les pauses que realitza un emissor en la lectura de textos. Es pretén, llavors, trobar una correlació entre les dades extretes del corpus analitzat, el context d'aparició dels signes de puntuació i els parlants per tal de aportar possibles millores a les aplicacions de creació i recreació de text com les descrites a l'apartat 2.2.2.

4.2 Metodologia

4.2.1 Corpus

El corpus emprat en l'experiment consisteix en 3 texts (vegeu l'annex 2A) que contenen 60 comes (classificades segons els criteris exposats a l'apartat 3.2) i 24 punts i seguit. Les pauses estudiades es troben entre paraules d'una, dos, tres i quatre síl·labes, sempre seguint l'esquema presentat a l'annex 2B.

En el cas de les consonants que segueixen una pausa no s'han considerat els sons oclusius sords, ja que aquesta classe de consonants presenta una fase de tancament en la qual la freqüència fonamental és zero. Conseqüentment amb aquest fet, els resultats sobre la durada de les pauses silencioses es veurien sensiblement afectats i contindrien un error de mesura més gran.

Les gravacions duraven aproximadament 5 minuts per locutor. En total s'han analitzat 20 minuts de lectures amb un total de 240 comes i 96 punts. La taula 1 mostra el total de casos per a cada categoria analitzada.

	Vocal – Pausa – Vocal	Vocal – Pausa – Consonant
Comes d'enumeració	40	40
Comes amb oracions subordinades relatives	40	40
Comes amb construccions adverbials	40	40
Nombre total de comes	120	120
Punts i seguit	48	48
Nombre total de signes de puntuació	168	168

Taula 1. Número i tipus de signes de puntuació analitzats.

4.2.2 Locutors

Per a la realització d'aquest estudi, s'han seleccionat quatre locutors catalanoparlants, dos masculins i dos femenins, d'entre 23 i 25 anys amb estudis universitaris pràcticament finalitzats. S'ha procurat que tots els parlants tinguessin el mateix domini lingüístic del català.

4.2.3 Gravació

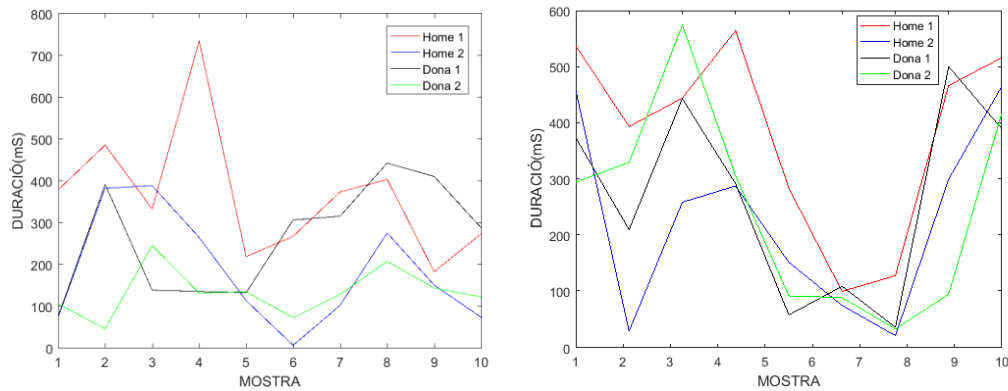
La gravació s'ha realitzat a les cabines d'assaig que l'Ajuntament de Ripollet facilita als músics del poble. La sala facilitada compta amb un bon aïllament sonor de l'exterior i, d'entre el material possible per a la gravació del corpus, es va escollir un micròfon Rode NT1-A amb *antipop* i amb una resposta en freqüència de 20 Hz a 20 kHz. Les gravacions van ser recollides amb un MacBook Pro i5 2,0 GHz 256 GB mitjançant el programa Cubase Pro (versió 8) i els arxius utilitzats per a l'experiment van ser guardats amb l'extensió .wav a una freqüència de mostreig de 44100 Hz. Finalment, la connexió entre el micròfon i l'ordinador es va fer a través d'un preamplificador de micròfon Behringer U-Phoria UM2.

4.2.4 Procediment d'anàlisi

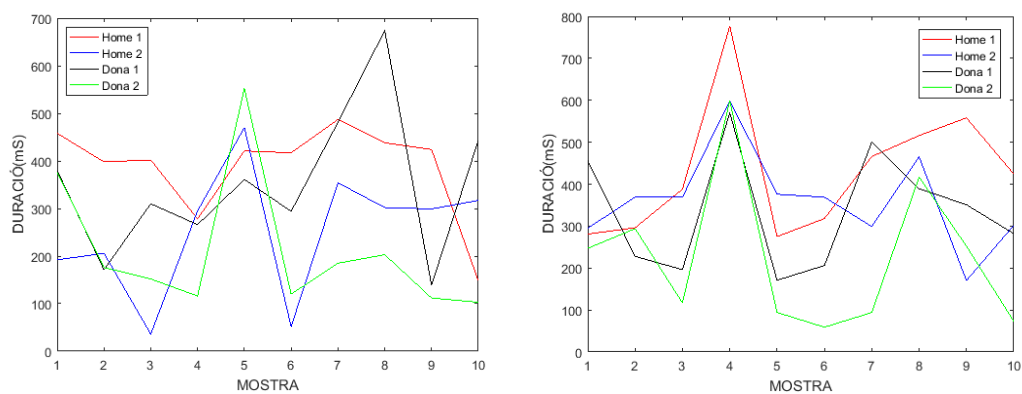
La durada de les pauses s'ha analitzat amb el programa Audacity (versió 2.1.2). En el càlcul dels valors mitjans de cada tipus de pausa, no s'han tingut en compte tots aquells valors inferiors als 200 ms, tal i com s'ha establert a l'apartat 3.1. Per a l'estudi i la realització de les funcions gràfiques s'ha fet servir el programa MatLab (versió R2016b), amb els codis de programació facilitats en l'annex 3. A les figures de l'apartat 4.3 que recullen els resultats es presenta, per a cada context d'aparició dels signes de puntuació estudiats, la durada de les pauses en mil·lisegons a l'eix vertical i el lloc en el qual apareixen en el text a l'eix horitzontal (indicat com a 'mostra').

4.3 Resultats

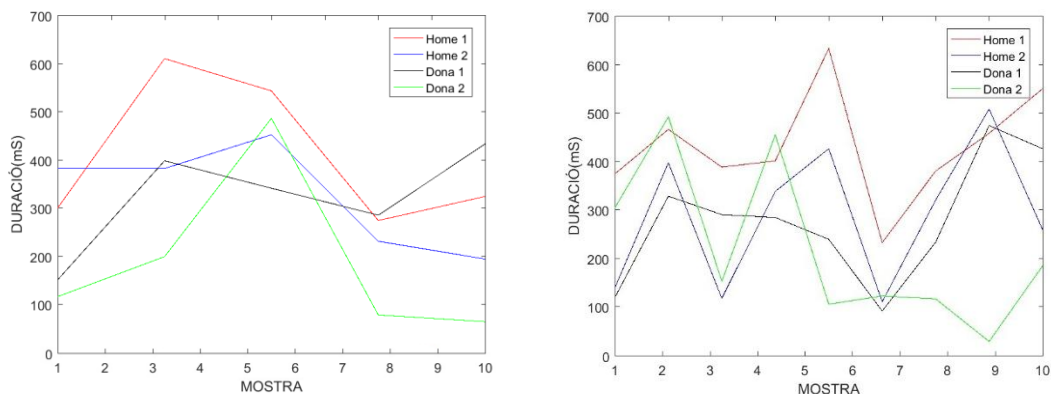
Els resultats que es mostren a les taules 2 i 3 de l'annex 1A i a les figures 5, 6, 7, 8, 9 i 10 posen de manifest que existeix una relació entre la duració dels diferents tipus de pauses estudiades i els parlants. Tot i seguir el mateix patró de comportament, els valors obtinguts generalment són força diferents entre parlants; és a dir, es pot observar una tendència clara a representar la mateixa funció, tot i que desplaçada en l'eix vertical o horitzontal segons el tipus de pausa (comes d'enumeració, amb oracions subordinades de relatiu, amb construccions adverbials i punts i seguits) i el parlant.



Figures 5 i 6. Comportament de les comes d'enumeració
(a l'esquerra: vocal-pausa-vocal; a la dreta, vocal-pausa-consonant).



Figures 7 i 8. Comportament de les comes amb oracions subordinades de relatiu
(a l'esquerra: vocal-pausa-vocal; a la dreta, vocal-pausa-consonant).



Figures 9 i 10. Comportament de les comes amb construccions adverbials
(a l'esquerra: vocal-pausa-vocal; a la dreta, vocal-pausa-consonant).

Es pot establir que la durada de la pausa corresponent a una coma, en el corpus recollit per a aquest experiment, oscil·la entre els 100 i els 777 ms. La coma realitzada amb durada més breu s'ha trobat en una enumeració corresponent al context 'paraula acabada en vocal – pausa – paraula començada en consonant no oclusiva' i la coma

realitzada amb una durada més llarga es troba abans d'una oració subordinada de relatiu, també en el mateix context.

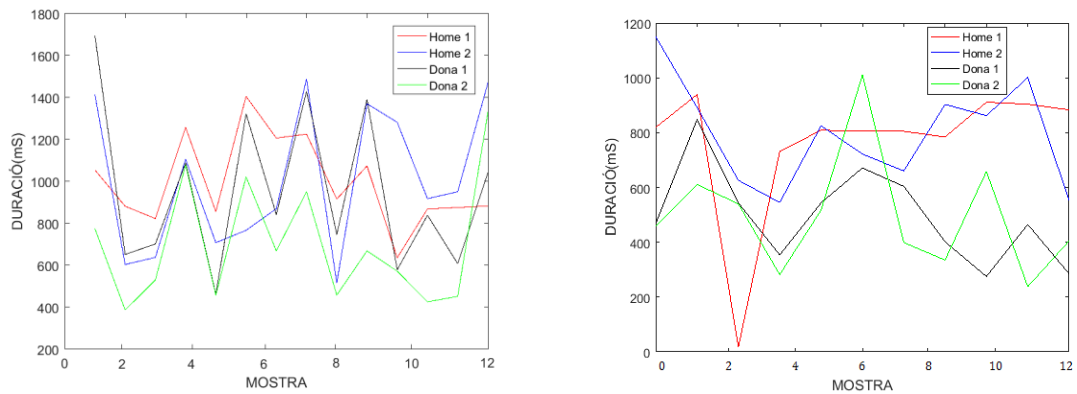
La tendència a que la durada d'una pausa silenciosa corresponent a una coma sigui més breu o més llarga depèn única i exclusivament del parlant que la realitzi. És cert que, tant els valors de les taules 2 i 3 de l'annex 1 com els valors recollits en les figures mostren que, en línies generals, les pauses més llargues corresponents a comes apareixen entre una paraula acabada en vocal i una altra que comença en consonant, però no sembla una característica determinant per poder afirmar que depenen, també, del context en el qual apareixen. La confirmació que aquesta diferència no és significativa s'obté mitjançant la realització de la prova del χ^2 de Pearson⁽⁷⁾, en la qual els valors obtinguts per a cada parlant individualment oscil·len entre $p = .2181$ i $p = .2534$ després d'haver analitzat totes les combinacions possibles. Per tant, demostra que no hi ha diferències significatives entre els diferents tipus de comes (d'enumeració, amb oracions subordinades de relatiu i amb construccions adverbials) que apareixen als contextos 'vocal – pausa – vocal' i 'vocal – pausa – consonant'.

Des del punt de vista de les representacions gràfiques, s'observa un comportament matemàtic similar en relació a la realització de les pauses referents a les comes. Tot i que els valors de les pauses són molt diferents entre sí, els parlants les efectuen d'una manera molt similar quan es troben en una mateixa posició en els texts que s'han fet servir a l'experiment. Per exemple, la diferència entre la pausa silenciosa de la posició 1 i la de la posició 2 de l'eix horitzontal és manté proporcionalment en tots els casos i en tots els parlants.

En el cas de la pausa corresponent a punt i seguit, en canvi, els intervals de duració engloben valors entre els 238 i els 1696 ms. i tornen a ser molt diferents entre els parlants. El punt i seguit de durada més curta es localitza entre una paraula acabada en vocal i una paraula començada per consonant, mentre que el punt i seguit de durada més llarga i, de fet, la pausa silenciosa de més duració, es troba al context 'paraula acabada en vocal – pausa – paraula començada en vocal'. Com ja passava anteriorment, la funció es comporta de manera molt similar entre els parlants, com mostren les figures 11 i 12. Sembla que es posa de manifest, en aquest cas, una tendència a realitzar les pauses

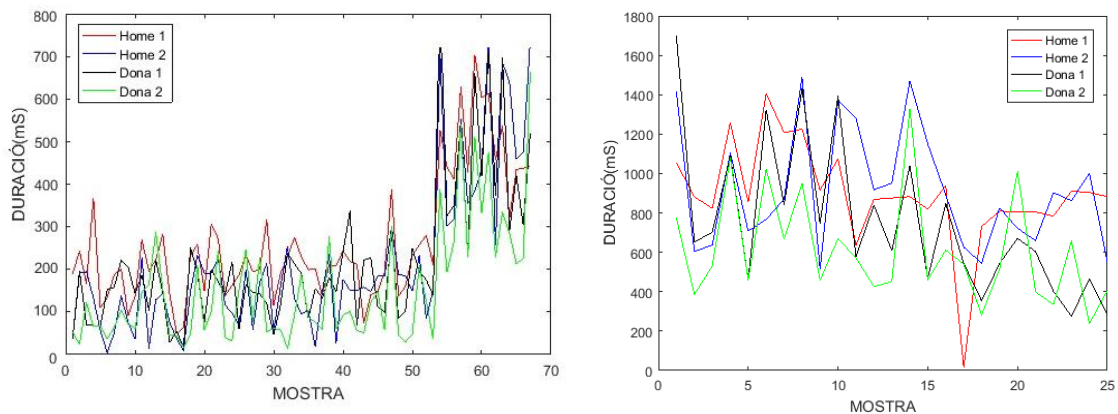
⁷Amb el test de χ^2 de Pearson es mesuren les diferències entre els valors de pauses silencioses del corpus i la seva distribució, per establir que la distinció entre aquests dos elements no s'ha realitzat de manera aleatòria.

corresponents als punts i seguit com a més llargues quan es troben entre una paraula que acaba en vocal i una que comença per vocal. També es manté, com en el cas de les pauses referents a les comes, la diferència proporcional entre les distàncies de dos punts de posició.



Figures 11 i 12. Comportament dels punts i seguit
(a l'esquerra: vocal-pausa-vocal; a la dreta, vocal-pausa-consonant).

Finalment, a les figures 13 i 14 es pot observar la funció de totes les pauses segons les mesures recollides a les taules 2 i 3.



Figures 13 i 14. Comportament de totes les pauses corresponents a
una coma (esquerra) i a un punt i seguit (dreta).

Sembla, llavors, que els parlants expressen les comes de manera diferent als punts i seguit, amb uns valors de durada més alts per aquests darrers signes de puntuació. Com mostra la figura 14, la duració màxima dels punts i seguit arriba a ser el doble dels valors màxims de duració de les comes. Aquest fet permetria discriminar les pauses silencioses en funció de si corresponen a comes o bé a punts i seguit, però aquesta

classificació no seria del tot encertada ja que es troben valors mínims de punts i seguit similars a valors normals de comes. És per aquesta raó que es considera necessari l'estudi de les variacions de la freqüència fonamental abans i després de les pauses per tal de complementar els resultats matemàtics que es presenten a totes les figures anteriors en forma de funció.

5. CONCLUSIONS

En els resultats de l'experiment sembla que existeix una forta correlació matemàtica entre els valors de les pauses silencioses estudiades i el seu context d'aparició. No s'ha trobat cap indici que demostrï que la distinció de gènere pugui incidir en els resultats. Els valors obtinguts per a les pauses silencioses corresponents a les comes i els punts i seguit en català es troben dins l'interval dels 200 i els 1696 ms, en el rang del que Campione i Véronis (2002) classifiquen com a “pauses mitjanes”. Cal afegir que, degut a les limitacions de temps i al fet de no disposar de corpus orals de domini públic adequats als objectius de l'estudi, només s'ha pogut realitzar una gravació per a cada parlant. Seria òptim dur a terme el mateix experiment amb un corpus més ampli i amb més d'una lectura per a cada parlant per tal d'analitzar la variabilitat intra-locutor en la durada de les pauses silencioses relacionades amb els signes de puntuació.

Aquest experiment es podria continuar estudiant els canvis que es donen abans i després de les pauses. Això seria interessant per poder definir, mitjançant una equació, els valors de durada dels signes de puntuació. Amb les equacions definides es podria aconseguir que una aplicació de creació i recreació de text fos capaç de calcular automàticament els valors de durada d'una pausa d'un parlant determinat i situar-los en el marc del règim transitori del senyal sonor. El règim transitori d'un senyal elèctric és aquell estat inestable que té lloc just abans de començar a funcionar normalment i després d'apagar l'aparell. En el cas del senyal sonor, el règim transitori estaria delimitat justament abans d'interrompre el discurs oral i, també, justament després de la interrupció. Descriuint matemàticament aquesta particularitat, les pauses silencioses estarien definides, a més de pels seus valors en mil·lisegons, per una equació polinòmica de segon grau o superior que reduiria molt significativament els intervals actuals de pauses silencioses, fet que representaria tota una renovació en les aplicacions de creació i recreació de text.

Finalment, una possible aplicació de l'experiment que s'ha presentat consistiria en la realització d'un subprograma que permetés a una aplicació de creació i recreació del text mostrar unes oracions que l'usuari llegís en veu alta a fi de realitzar una "calibració" dels intervals de duració de les pauses silencioses del parlant i fer possible, així, l'assignació automàtica de signes de puntuació.

REFERÈNCIES BIBLIOGRÀFIQUES

- ALBALÁ, M. J., BATTANER, E., CARRANZA, M., DE LA MOTA, C., GIL, J., LLISTERRI, J. i RÍOS, A. (2008): «VILE: Análisis estadístico de los parámetros relacionados con el grupo de entonación», a Pamies, A., Amorós, M. C. i Pazos, J. M. (Eds.), *Language Design. Journal of Theoretical and Experimental Linguistics. Special issue 2: Experimental prosody*, (pp 15-22). Granada: Método Ediciones.
- BAKER, P., HARDIE, A., i McENERY, T. (2006): «A glossary of corpus linguistics». Edinburgh: Edinburgh University Press.
- BRENNAN, S. i WILLIAMS, M. (1995): «The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers» *Journal of Memory and Language*, 34, (pp 383-398).
- BUTTERWORTH, B. (1980): «Evidences for pauses in speech» a *Language production, Volume 1: Speech and talk* (pp 155-176). London: London Academic Press.
- CAMPIONE, E. i VÉRONIS, J. (2002): «A large-scale multilingual study of pause duration» a *Speech Prosody 2002. Proceedings of the 1st International Conference on Speech Prosody*, (pp 199-202). Aix-en-Provence, França, 11-13 d'abril del 2002.
- CALSAMIGLIA, H. i TUSÓN, A. (1999): «Las cosas del decir. Manual de análisis del discurso». Barcelona: Ariel (2a edició, 2007).
- DUEZ, D. (1982): «Silent and non-silent pauses in three speech styles» *Language and Speech*, 25, (pp 11-28).
- DUXANS, H. i RUIZ, M. (2012): «Reconocimiento automático del habla». Barcelona: Universitat Oberta de Catalunya.
- GIL, J. (2007): «Fonética para profesores de español: De la teoría a la práctica». Madrid: Arco/Libros.
- HOUCQUE, D. (2005): «MatLab for engineering students». Illinois: Northwestern University.
- JUANG, B-H. i RABINER, L.R (1993): «Fundamentals of speech recognition». Englewood Cliffs, NJ: Prentice Hall.
- KENNEDY, G. (1998): «An introduction to corpus linguistics». London & New York: Longman.
- LEA, W.A. (1986): «Electronic speech recognition. Techniques, technology and applications». London: Collins.
- MACHUCA, M. J., LLISTERRI, J. i RÍOS, A. (2015): «Las pausas sonoras y los alargamientos en español: un estudio preliminar». *Normas. Revista de Estudios Lingüísticos Hispánicos*, 5, (pp. 81-96).
- MARIÑO, J. B.: «Reconocimiento de habla fluida (I)». *Curso sobre tecnologías del habla*. Grup de Tractament de la Parla, Departament de Teoria del Senyal i Comunicacions, Universitat Politècnica de Catalunya.

MARTÍ, M.A. i ALTRES (2001): «Les tecnologies del llenguatge». Barcelona: Universitat Oberta de Catalunya.

MÉNDEZ, J., HERNANDO, J., NADEU, C. i VALLVERDÚ, C. (1995): «Esquema unificado de parametrización de la señal de voz en reconocimiento del habla» a *X Simposim Nacional de la Unión Científica Internacional de Radio*. Valladolid, 27-29 de setembre de 1995. Libro de actas (pp 97-100). Valladolid: Universidad de Valladolid.

MIÑANA, L. (2014): «Una tirada de dados: Algunos hitos de la literatura digital y sus antecedentes» a *Presentación en II y IV Salón de Literatura Transmedia*. Etopía, Zaragoza, diciembre de 2014. (pp. 86-91).

MORENO, A. (1998): «Lingüística computacional». Madrid: Síntesis.

MOURE, T. i LLISTERRI, J. (1996): «Lenguaje y nuevas tecnologías. El campo de la lingüística computacional.» a Fernández Pérez, M. (Eds.), *Avances en lingüística aplicada* (pp. 147-228). Santiago de Compostela: Servicio de Publicación e Intercambio Científico.

NEWS, A. [Arirang News]. (26 d'Octubre de 2016): «Police arrest suspect in Washington shopping mall shooting; 8 injured in Baltimore gun attack»: [en línia]: <<https://www.youtube.com/user/arirangnews>>.

OOSTDIJK, N. i DE HAAN, P. (1994): «Corpus-based research into language». Amsterdam-Atlanta: Rodopi.

RABINER, L.R. (1989): «A tutorial on Hidden Markov Models and selected applications in speech recognition». *Proceedings of the IEEE*, 77, (pp 257-286).

REBOLLO, L. (1997): «Pausas y ritmo en el lenguaje oral. Didáctica de la pronunciación» a Moreno Fernández, F., Gil Bürmann, M. i Alonso, K. (Eds.), *El español como lengua extranjera: del pasado al futuro*. Actas del VIII Congreso Internacional de la Asociación para la Enseñanza del Español como Lengua Extranjera. Alcalá de Henares, 17 - 20 de septiembre de 1997, (pp 667-676). Alcalá de Henares: Servicio de Publicaciones de la Universidad de Alcalá.

RODRÍGUEZ FONOLLOSA, J.A. (2010): «La tecnologia de la parla en català. Avenços i reptes». *Llengua i ús. Revista tècnica de política lingüística*, 48, (pp 59-64).

ROSE, R. L. (1998): «The communicative value of filled pauses in spontaneous speech». Birmingham: University of Birmingham.

STEJSKAL, V., BOURBAKIS, N. i ESPOSITO, A. (2010): «Empty speech pause detection algorithms' comparison». *International Journal of Advanced Intelligence*, 2, 1, (pp. 145-160). República Txeca: University of Technology.

UNIVERSITAT AUTÒNOMA DE BARCELONA (1995): «Els signes de puntuació». Bellaterra: Gabinet de Llengua Catalana. Universitat Autònoma de Barcelona.

WAIBEL, A. (1986): «Suprasegmentals in very large vocabulary word recognition» a Schwab, E.E. i Nusbaum, H (eds.) *Pattern recognition by humans and machines. Volume I: Speech perception* (pp 159-186). Orlando: Academic Press.

ANNEX 1: Resultats per locutor

Annex 1A: Taules de resultats per parlant, context i signes de puntuació

	Parlant masculí 1	Parlant masculí 2	Parlant femení 1	Parlant femení 2
Comes d'enumeració (Vocal-pausa-Vocal)	379ms	74ms	76ms	106ms
	486ms	383ms	392ms	47ms
	333ms	389ms	139ms	246ms
	735ms	265ms	136ms	132ms
	219ms	114ms	133ms	135ms
	268ms	70ms	307ms	73ms
	374ms	103ms	316ms	129ms
	404ms	276ms	443ms	208ms
	183ms	151ms	411ms	144ms
	274ms	74ms	288ms	123ms
Comes d'enumeració (Vocal-pausa-Consonant)	537ms	457ms	373ms	294ms
	393ms	28ms	209ms	329ms
	444ms	258ms	443ms	574ms
	564ms	287ms	291ms	305ms
	283ms	151ms	57ms	91ms
	100ms	74ms	108ms	88ms
	127ms	20ms	35ms	32ms
	326ms	299ms	500ms	94ms
	466ms	354ms	424ms	265ms
	516ms	465ms	389ms	417ms
Comes amb oracions subordinades de relatiu (Vocal-pausa-Vocal)	458ms	192ms	380ms	376ms
	398ms	206ms	171ms	176ms
	402ms	36ms	310ms	152ms
	278ms	295ms	266ms	116ms
	421ms	470ms	361ms	552ms
	417ms	52ms	294ms	120ms
	487ms	354ms	481ms	185ms
	438ms	302ms	674ms	203ms
	424ms	299ms	139ms	112ms
	148ms	317ms	443ms	103ms
Comes amb oracions subordinades de relatiu (Vocal-pausa-Consonant)	281ms	295ms	456ms	247ms
	296ms	369ms	228ms	294ms
	388ms	599ms	196ms	117ms
	777ms	376ms	570ms	599ms
	275ms	369ms	171ms	94ms
	318ms	59ms	206ms	59ms
	92ms	299ms	342ms	94ms
	466ms	465ms	500ms	417ms
	558ms	170ms	389ms	252ms
	424ms	302ms	282ms	73ms

Taula 2. Valors mesurats de les comes d'enumeració i de les que delimiten oracions subordinades de relatiu.

	Parlant masculí 1	Parlant masculí 2	Parlant femení 1	Parlant femení 2
Comes amb construccions adverbials (Vocal-pausa-Vocal)	301ms	383ms	152ms	117ms
	611ms	383ms	399ms	200ms
	544ms	453ms	297ms	487ms
	275ms	395ms	256ms	365ms
	624ms	437ms	187ms	287ms
	586ms	359ms	265ms	325ms
	436ms	267ms	307ms	216ms
	256ms	232ms	342ms	134ms
	275ms	216ms	286ms	79ms
	325ms	195ms	434ms	65ms
Comes amb construccions adverbials (Vocal-pausa-Consonant)	375ms	140ms	120ms	305ms
	467ms	398ms	329ms	493ms
	389ms	118ms	291ms	153ms
	402ms	339ms	285ms	456ms
	634ms	427ms	240ms	106ms
	139ms	250ms	124ms	59ms
	233ms	111ms	92ms	123ms
	381ms	321ms	234ms	117ms
	459ms	509ms	475ms	29ms
	551ms	260ms	427ms	185ms
Punts i seguit (Vocal-pausa-Vocal)	1055ms	1415ms	1696ms	775ms
	882ms	604ms	652ms	387ms
	823ms	638ms	702ms	530ms
	1258ms	1106ms	1089ms	1079ms
	856ms	708ms	462ms	458ms
	1406ms	767ms	1323ms	1022ms
	1207ms	870ms	842ms	669ms
	1226ms	1489ms	1430ms	951ms
	916ms	516ms	747ms	458ms
	1074ms	1371ms	1392ms	669ms
	636ms	1282ms	579ms	572ms
	869ms	918ms	839ms	426ms
Punts i seguit (Vocal-pausa-Consonant)	819ms	1150ms	468ms	460ms
	939ms	892ms	848ms	611ms
	18ms	626ms	544ms	540ms
	731ms	545ms	354ms	282ms
	809ms	825ms	544ms	517ms
	805ms	722ms	671ms	1010ms
	805ms	660ms	604ms	399ms
	784ms	903ms	402ms	335ms
	911ms	862ms	275ms	658ms
	904ms	1001ms	465ms	238ms
	876ms	951ms	608ms	452ms
	883ms	553ms	285ms	405ms

Taula 3. Valors mesurats de les comes amb construccions adverbials i dels punts i seguit.

Annex 1B: Durada mitjana de les pauses en funció del signe de puntuació i del parlant

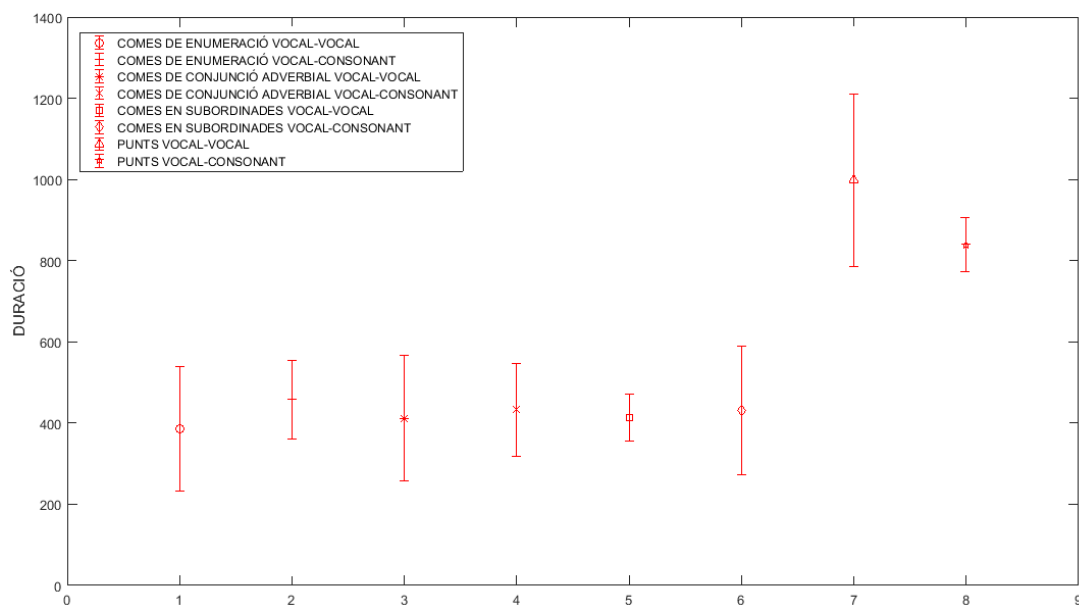


Figura 15. Durada mitjana (en ms) i desviació estàndard de les pauses en funció del signe de puntuació del parlant masculí 1.

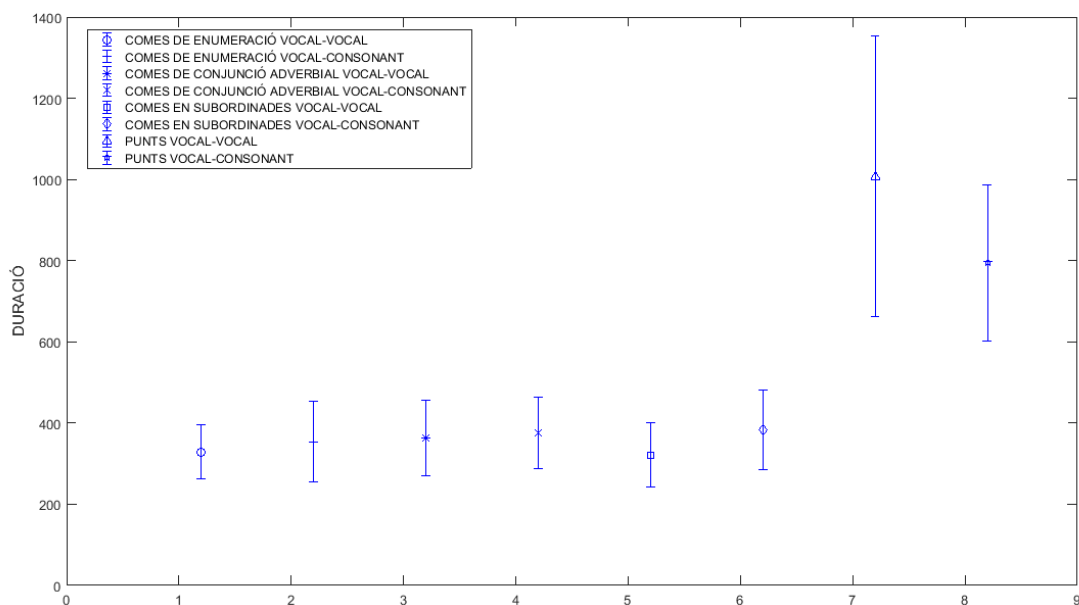


Figura 16. Durada mitjana (en ms) i desviació estàndard de les pauses en funció del signe de puntuació del parlant masculí 2.

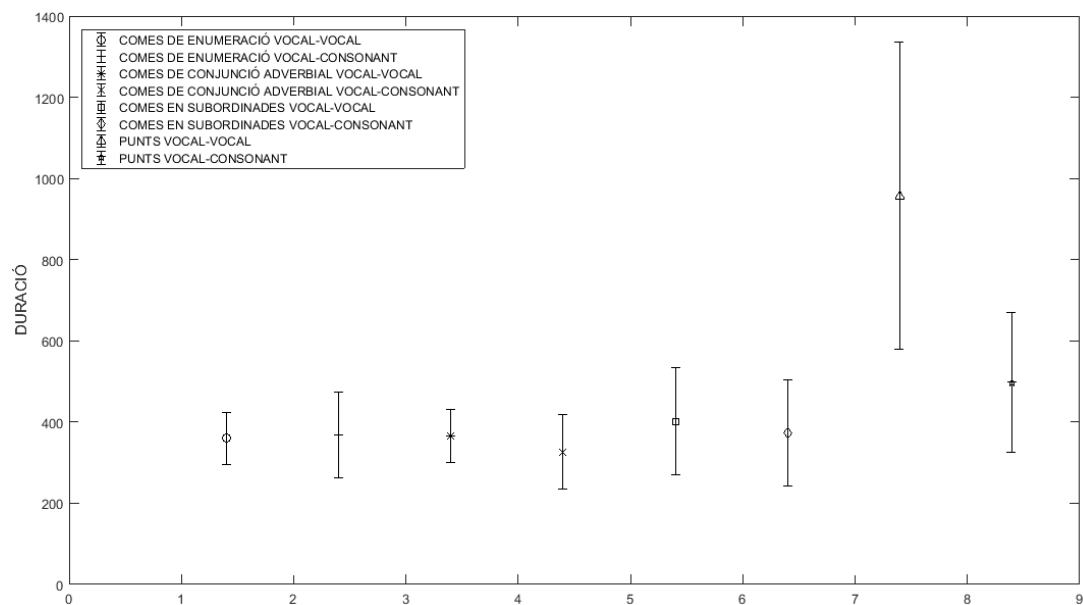


Figura 17. Durada mitjana (en ms) i desviació estàndard de les pauses en funció del signe de puntuació del parlant femení 1.

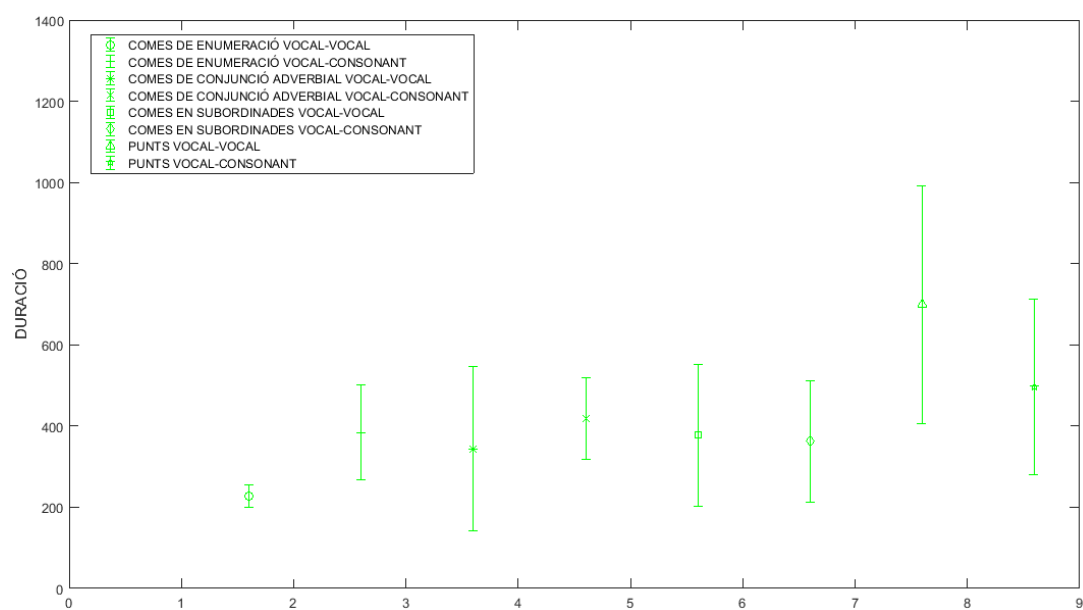


Figura 18. Durada mitjana (en ms) i desviació estàndard de les pauses en funció del signe de puntuació del parlant femení 2.

Annex 1C: Durada mitjana de les pauses en funció del context, del signe de puntuació i del parlant

	Parlant masculí 1	Parlant masculí 2	Parlant femení 1	Parlant femení 2
Comes d'enumeració (Vocal-pausa-Vocal)	385,77ms	328,25ms	359,50ms	227ms
Comes d'enumeració (Vocal-pausa-Consonant)	457,57ms	353,20ms	367,50ms	383,80ms
Comes amb oracions subordinades de relatiu (Vocal-pausa-Vocal)	413,66ms	320,42ms	401,12ms	377ms
Comes amb oracions subordinades de relatiu (Vocal-pausa-Consonant)	429,90ms	382,55ms	372,75ms	361,80ms
Comes amb construccions adverbials (Vocal-pausa-Vocal)	411,20ms	362,75ms	365,25ms	343,50ms
Comes amb construccions adverbials (Vocal-pausa-Consonant)	432,33ms	375,66ms	325,85ms	418ms
Punts i seguit (Vocal-pausa-Vocal)	997,64ms	1007,5ms	957,28ms	698,42ms
Punts i seguit (Vocal-pausa-Consonant)	839ms	794,45ms	496,36ms	495,90ms

Taula 4. Durada mitjana de les pauses corresponents a les comes i als punts i seguit segons el context.

ANNEX 2: Corpus i contexts d'aparició de signes de puntuació

Annex 2A: Corpus de textos

TEXT A

El poema “Els cocoters de Macuto” pertany al llibre *El cor quiet* (1925) de Josep Carner, el poeta més significatiu del Noucentisme. La seva obra segueix plenament el concepte noucentista, com ara la descripció d’una naturalesa civilitzada, la creació literària entesa com un rigorós procés d’elaboració i la voluntat d’aconseguir un model de llengua moderna. Josep Carner és considerat un renovador de la llengua literària. Aquest poema està inclòs a la secció “L’arbre”. En totes les composicions d’aquesta secció, les quals són quatre, l’arbre (eucaliptus, ametller, pi, pollancre...) són el motiu del poema, el pretext per desenvolupar altres temes i la manera d’expressió del poeta.

El poema està format per sis quartets de versos heptasíl·labs i la rima és assonant amb el tercer vers lliure: abcb, defe...; i alternen els vers femení amb el masculí. L’última estrofa apareix visualment dividida, tot i que mètricament està lligada. Aquesta separació vol donar èmfasi al valor dels punts suspensius que tanquen els dos versos del principi. El poema parteix d’una escena que ha quedat fixada en la imaginació, en el record i el somni del poeta: quatre cocoters davant del mar que va veure a la ciutat de Macuto, Veneçuela. El poeta durant la seva carrera diplomàtica s’estigué a Amèrica una temporada. D’aquesta manera, el tema del poema és un detall intranscendent de la realitat. Però, de debò, la descripció d’aquesta imatge conservada en el record permet expressar les sensacions que li va produir a contemplació d’aquella visió. El poema es pot seccionar en quatre, segons la seva intenció. En la primera estrofa s’introdueix l’escena, la qual descriu, de manera objectiva, uns cocoters a Macuto davant el mar i remarca la casualitat del fet per un present de l’atzar.

Les dues estrofes següents desenvolupen la descripció, ara ja a través de la metàfora, de l’arbre. La quarta i la cinquena estrofa continuen amb les metàfores que ara fan referència al creixement que han fet els arbres d’abans fins ara. Com a conclusió, en l’última quarteta apareix el jo, que deixa de banda la realitat externa del paisatge i introdueix la reflexió, els seus pensaments i la seva percepció.

Font: Comentari literari: Josep Carner. “Els cocoters de Macuto”, a *Literatura catalana*.

Marta, aprova’m. <http://literaturacatalanamarta.blogspot.com.es/p/comentari-detext>.

TEXT B

Els senyors Banyó, Elía, Ganiveta, Garcia, Imago, Silveri i Valls ja han arribat a l'acte. Les dones de cadascú, com és obvi, també han vingut com acompanyants; cadascuna per donar suport al seu home. Com passa en cada acte de graduació, els professors encarregats del discurs lluiten per ser aquell que arribi al cor de les famílies i dels estudiants que, d'aquí poca estona, seran obsequiats amb un diploma. Quin goig que fa veure'ls tan contents!

Tot i així, sempre n'hi ha alguna, de discrepància: que si l'hora de celebració no és del tot adequada, que si les invitacions per persona no es reparteixen de manera equitativa, que si hi ha massa gent a l'auditori... Però, fora de les ja típiques amonestacions relacionades amb l'acte, l'ambient que es respira és extremadament emotiu. Un cop acabada la primera lectura, els alumnes Adina, Aguilà, Andreu, Aria, Aseta, Ateu i Aznar pugen a l'escenari per tal de saludar el rector i el director, així com per rebre el diploma. Després d'aquesta tradició, els alumnes s'ajunten per fer-se una foto de grup alhora que somriuen amb els ulls plens d'energia. Aquesta experiència l'aniran sentint tots els alumnes posteriors i s'anirà interpel·lant amb el discurs propi de cada professor participant a l'acte. El final, com no podria ser de cap altra manera, estaria protagonitzat pels alumnes el cognom dels quals comencés per la zeta.

Un cop el responsable del discurs finalitza, tothom surt de la sala adequadament i el rebombori inunda els passadissos de l'edifici. La festa, el més probable, és que continuï per a molts d'ells.

Font: Elaboració pròpia

TEXT C

Els llums seguien apagats. El gos no bordava i jo, tan innocent com sempre, esperava que la porta s'obris d'aquí poc. Pensava que, ja que el timbre no funcionava, el meu germà posaria més atenció quan truqués a la porta però, en lloc d'això, podia sentir la música de Pearl Jam que sonava a tot volum per la casa i s'escampava fins arribar a la porta.

Anava amb aquella suadora que tant li agradava: aquella verda, amb uns bolets de temporada, un punt blau al mig del pit i una frase que deia “`principi del final”. [...] I jo continu aquí fora, asseguda al porxo i separada de la música, el meu germà i el meu gos. Fora: lluny d'aquella felicitat tan pragmàtica, aïllada, salvatge i elemental. Lluny de la zona de confort però, de la mateixa manera, molt a prop de tot allò que l'envoltava: El sol de llum apagada, l'aire de mitja tarda, els cotxes que estacionaven amb timidesa i el rerefons de la cançó Even Flow. I es que, tot i passar dels trenta, els somnis d'infantesa del meu germà encara eren massa vius.

Font: Elaboració pròpia

Annex 2B: Contexts d'aparició dels signes de puntuació

PUNT I SEGUIT (24):

-Paraula acabada en vocal – pausa – Paraula que comença amb consonant NO OCLUSIVA (12).

- Paraula acabada en vocal – pausa – Paraula que comença amb vocal (12).

COMA (60):

- ENUMERACIÓ (20)

-Paraula acabada en vocal – pausa – Paraula que comença amb consonant NO oclusiva (10)

- Paraula acabada en vocal – pausa – Paraula que comença amb vocal (10)

- OR. EXPL. REL. (20)

-Paraula acabada en vocal – pausa – Paraula que comença amb consonant NO oclusiva (10)

-Paraula acabada en vocal – pausa – Paraula que comença amb vocal NO oclusiva (10)

- CONST. ADVERBIALS (20)

-Paraula acabada en vocal – pausa – Paraula que comença amb consonant (10)

-Paraula acabada en vocal – pausa – Paraula que comença amb vocal (10)

TOTAL: 240 comes (60 x 4 parlants) i 96 punts i seguit (24 x 4 parlants)

ANNEX 3: Codis de *MatLab* per a la realització de les figures

```

clear all;close all; % resetejar les dades
%%
Data = load ('ENUM.VOCVOC.txt'); % carregar les dades
% treure els valors menors de 200ms
DataData=Data;
DataData1=DataData(:,1);
DataData2=DataData(:,2);
DataData3=DataData(:,3);
DataData4=DataData(:,4);
Index1 = find(DataData1<200);
DataData1(Index1)=[];
Index2 = find(DataData2<200);
DataData2(Index2)=[];
Index3 = find(DataData3<200);
DataData3(Index3)=[];
Index4 = find(DataData4<200);
DataData4(Index4)=[];

% mitjana i desviació estandard
MEDIA11= mean (DataData1);
STD11 = std(DataData1);
MEDIA12= mean (DataData2);
STD12 = std(DataData2);
MEDIA13= mean (DataData3);
STD13 = std(DataData3);
MEDIA14= mean (DataData4);
STD14 = std(DataData4);
% borrar les variables temporals
clearvars DataData DataData1 DataData2 DataData3 DataData4 Index1 Index2 Index3 Index4
% creació de figures
figure(1)
plot (Data(:,1),'r')
hold on
plot (Data(:,2),'b')
hold on
plot (Data(:,3),'k')
hold on
plot (Data(:,4),'g')
%title('ESTUDI DE COMES DE ENUMERACIÓ ENTRE VOCAL I VOCAL')
xlabel('MOSTRA')
ylabel('DURACIÓ(ms)')
legend('Home 1','Home 2','Dona 1','Dona 2')
%% CONCATENACIÓ %
% les figures es fan de la mateixa manera que anteriorment.
% totes les comes
Data9=[Data; Data2; Data3 ;Data4 ;Data5 ;Data6 ;Data7];
% tots els punts
Data10 =[Data7;Data8];

```

```

%% figura de les mitjanes de cada parlant
figure(11)
errorbar(1,MEDIA11,STD11,'or')
hold on
errorbar(2,MEDIA21,STD21,'+r')
hold on
errorbar(3,MEDIA31,STD31,'*r')
hold on
errorbar(4,MEDIA41,STD41,'xr')
hold on
errorbar(5,MEDIA51,STD51,'sr')
hold on
errorbar(6,MEDIA61,STD61,'dr')
hold on
errorbar(7,MEDIA71,STD71,'^r')
hold on
errorbar(8,MEDIA81,STD81,'pr')
xlim ([0 9]);
ylim ([-1, 1400 ]);
ylabel('DURACIÓ');
%title('HOME 1')
legend ('COMES DE ENUMERACIÓ VOCAL-VOCAL','COMES DE ENUMERACIÓ VOCAL-
CONSONANT','COMES DE CONJUNCIÓ ADVERBIAL VOCAL-VOCAL', 'COMES DE
CONJUNCIÓ ADVERBIAL VOCAL-CONSONANT','COMES EN SUBORDINADES VOCAL-
VOCAL','COMES EN SUBORDINADES VOCAL-CONSONANT','PUNTS VOCAL-
VOCAL','PUNTS VOCAL-CONSONANT');
%% fem el mateix per les dades dels altres 3 parlants

%% Test estadistic %% chi2

[tbl,chi2stat,pval] = crosstab(Data(:,1),Data2(:,1));
[tbl,chi2stat,pval1] = crosstab(Data(:,2),Data2(:,2));
[tbl,chi2stat,pval2] = crosstab(Data(:,3),Data2(:,3));
[tbl,chi2stat,pval3] = crosstab(Data(:,4),Data2(:,4));
% guardar els valors
pvalEnum = [pval, pval1, pval2,pval3];
clearvars pval pval1 pval2 pval3
% Fem el mateix per els diferents conjunts de dades

```